DOCUMENT RESUME

ED 284 884 TM 870 427

AUTHOR Olejnik, Stephen
TITLE Teacher Education Effects: Looking Beyond the Means.
PUB DATE Apr 87
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Analysis of Variance; Bachelors Degrees; *Classroom Research; *Educational Attainment; Elementary Secondary Education; *Individual Differences; Masters Degrees; Mathematical Models; Predictor Variables; Simulation; Student Characteristics; *Teacher Behavior; *Teacher Education

ABSTRACT

Research investigations on teacher training effects have focused almost exclusively on differences between group means. The present paper suggests that several interesting and important research questions might be answered by examining student variability both within and between classrooms. Student variability might be considered as an outcome to be studied when teacher behavior affects heterogeneity or as a nuisance variable that should be controlled. A hypothetical data set is presented to demonstrate valid statistical methods which could be used in investigations of teacher training effects. (Author)

ED284884

Teacher Education Effects: Looking Beyond the Means

Stephen Olejnik

University of Georgia

Paper presented at the American Educational Research Association meetings in Washington D.C. April, 1987.

Abstract

Research investigations on teacher training effects have focused almost exclusively on differences between group means. The present paper suggests that several interesting and important research questions might be answered by examining student variability both within and between classrooms. Student variability might be considered as an outcome to be studied when teacher behavior affects heterogeneity or as a nuisance variable that should be controlled. A hypothetical data set is presented to demonstrate valid statistical methods which could be used in investigations of teacher training effects.

Teacher Education Effects: Looking Beyond the Means

Introduction

Most teachers would agree that what goes on in a classroom depends a great deal on the students within the class. How effective a teacher is in terms of increasing student achievement depends to some extent on the ability levels of the students. Research studies on teacher education effects generally (but not always) have taken into consideration student ability levels or previous achievement when comparing teacher training programs (e.g. Davis, 1964; Winkler, 1975; Murnane, 1975; Summers and Wolfe 1975; Ashton, Crocker and Olejnik 1986). But beyond student ability levels, classrooms can also differ regarding the variability of students within the class. This variability may be a function of pre-existing individual differences or may be a function of teacher behavior. When variability is the result of teacher behavior, dispersion might be an outcome to be studied and when variability is the result of pre-existing individual differences then variability might be considered as a nuisance variable to be controlled. In either case, teacher education effects research has given very little attention to classroom variability either within classrooms or between classrooms. The present paper considers dispersion as a variable to be studied and/or controlled. Using a hypothetical data set statistical

procedures are suggested and demonstrated which might be used in future studies on the effects of teacher training to investigate issues of variability.

Variability as an outcome

As an outcome measure variability can reflect the instructional philosophy or strategy adopted by a teacher. It can also reflect the consistency of behavior achieved by teachers using the same instructional strategy. Teacher behavior can affect the variability of student behavior within a classroom by focusing on different ends of the behavior distribution of interest. In the case of student achievement teacher behavior could reduce the variability of student achievement by focusing attention and instructional time on those students at the lower end of the achievement distribution. This instructional strategy might be adopted in response to minimum competence testing where the objective is to get as many students as possible to some predetermined achievement level. On the other hand teacher behavior could increase the variability of student achievement within a class by focusing attention at the higher end of the ability distribution and letting the lower levels fall farther and farther behind. This instructional strategy might be adopted where the objective is to maximize group gains. A teacher or school might adopt this strategy when merit pay is related to mean achievement gains by the students.

The consistency of achievement effects across classrooms or teachers can also be studied by examining the variability of outcomes. In the area of teacher training considerable research has indicated that mean classroom achievement does not depend on teacher training. These research studies have not examined the consistency of mean achievement gained by teachers having the same training or to compare the consistency of teachers having different training. Given that average achievement is similar, consistency of performance might be viewed very positively. Training programs which produce some excellent teachers and some poor teachers might be less desirable than a program which consistently produces good teachers.

Variability as a nuisance

It is generally recognized that in nonexperimental research studies statistical adjustments are often necessary to consider initial differences in achievement or ability. What is frequently ignored is that classrooms can also differ in the variability of achievement or ability levels within the classroom. And that this difference can affect what a teacher can do and what effect the teacher's behavior will have on the students. Teaching a class having students who differ greatly in their abilities is a considerably different experience from teaching a class in which students are very similar to each other in ability and background knowledge. Furthermore, classroom heterogeneity can have an effect whether classrooms have equal or unequal mean ability

levels. In studies of teacher training programs students are not randomly assigned to classes and classes are selected from different schools in a district and from different districts around a state. It seems likely that there will be differences in instructional philosophies at the school or district levels regarding the grouping of students within a classroom. It seems reasonable to think then that some differences in student variability within classes would be observed. Previous studies have ignored this issue and as a result may not have controlled a possible kev variable affecting teacher training effects.

Comparing classroom variability

If one considers within classroom variability as an outcome factor to be studied or as a nuisance factor to be controlled an importatant question of interest is whether some classes are more variable than others. One reason why researchers have ignored this question has been that good statistical procedures for comaparing variances have not been available. Tests of variance equalitv most often cited in statistics texts (eg. Bartlett, 1937; Cochran, 1941; Hartley 1950) are all extremely sensitive to the data distribution form. If the data have a non-normal distribution then these tests of variance equality can either overestimate or underestimate the probability of a Type I error. Although several alternative parametric and nonparametric tests have been suggested these procedures either do not control the Type I error rate or they lack

sufficient statistical power to be of practical value (Conover, W.J., Johnson, M.E. and Johnson, M.M., 1981). A procedure suggested by O'Brien (1978) however does provide a technique for comparing within classroom variability which is generally insensitive to the distributional form and has reasonable statistical power (Olejnik, S. and Algina, J., 1987). The procedure does underestimate the Type I error rate when the distribution form is leptokurtic and for those distributions it has lower statistical power. But most public school classrooms contain between 25 and 35 students which should be a more than sufficient sample size to compensate for the reduced statistical power associated with the distribution form.

O'Brien's procedure uses the analysis of variance model with the dependent variable created by transforming the original observations to a measure that reflects the group variability. Although several modifications of the transformation are possible, O'Brien (1981) recommends the following approach:

$$r_{ij} = \frac{[(n_j - 1.5)n_j(X_{ij} - \bar{X}_j)^2 - .5s_j^2(n_j - 1)]}{(n_j - 1)(n_j - 2)} \quad (1)$$

Where n_j is the number of observations in group j;

X_{ij} is the score for individual i in group j;

$\bar{X}_j$ is the mean of observations in group j;

and s_j^2 is the variance of observation in group j.

To demonstrate this procedure consider the hypothetical

data set reported in Table 1. Ten classrooms were randomly selected and 10 students were randomly selected within those

---

Insert Table 1 here

---

classes and given a 20 item test. In this problem classroom is a random factor rather than a fixed factor so the question that might be asked is whether classrooms varied with respect to the within group variability. The hypothesis can be written as $H_0: \sigma_c^2 = 0$. Table 2 presents the transformed scores using equation 1. It is

---

Insert Table 2 here

---

worth noting that the classroom means on the transformed variable are equal to the classroom variances of the original observations. Calculating the ANOVA F-ratio using the transformed variable resulted in a computed test statistic equalling 2.14. The critical test F statistic with 9 and 90 degrees of freedom is equal to 2.0 at the .05 level of significance, so there is sufficient evidence to reject the null hypothesis and conclude that classrooms do vary regarding the within classroom variability for this data set. If this difference in variability was viewed as a nuisance factor then it would be desirable to control it or

hold it constant when comparing the classrooms on other variables. Controlling within classroom variability is discussed in a later section of this paper. If variability is viewed as an outcome then the researcher might be interested in determining what variables are related to this difference in classroom variability.

Suppose the first five classrooms were taught by teachers having a master's degree and the last five classrooms were taught by teachers having a bachelor's degree. A researcher might ask the question whether the within classroom variability among teachers with a master's degree differ from the within classroom variability among teachers' with a bachelor's degree. The research design is hierarchical with classrooms nested in degree level of the teacher and students nested within classrooms. Both classrooms and students are random factors but the degree level of the teacher is a fixed factor. Table 3 reports the ANOVA summary table using the transformed scores as the dependent measure. The computed F statistic for the teacher

---

Insert Table 3 here

---

degree level was equal to 19.99 which exceeds the critical F statistic at the .05 level ($F_{1,8,.05}=5.32$). The within classroom variability was greater among teachers having a bachelor's degree than teachers having a master's degree.

This analysis is equivalent to the analysis conducted by Brown and Saks (1975) who examined the relationship between within classroom variability and several teacher background variables. The present analysis provides a second statistical test and answers an additional research question. The test for classrooms nested within degree program answers the following three related questions: (a) does the variability within classrooms vary among teachers having a master's degree; (b) does the variability within classrooms vary among teachers having a bachelor's degree and (c) does within classroom variability vary within both degree levels. The computed F-ratio for this statistical test equalled .688 which was less than the critical F statistic at the .05 level of significance ($F_{8,90,.05}=2.06$). It is concluded that the within classroom variability does not vary within either degree level.

To answer a question regarding the consistency of teacher training an analysis similar to that described above could be carried out. Typically teacher training effects researchers have asked whether students taught by teachers who were trained differently, differ in their average achievement. For example do students achieve at different levels if their teacher has a bachelor's or master's degree? Most of the research on teacher training effects do not provide sufficient evidence to indicate that average classroom achievement differ significantly (e.g. Katzman, 1971; Murnane, 1975; Summers and Wolfe, 1975; Brown and

Saks, 1975). Researcher's have not examined whether different training programs differed in the consistency of average student achievement. This could be answered by comparing the variability of mean classroom achievement of teachers from different training programs. Again using the data in Table 1 assume that the first five classrooms were taught by teachers with a master's degree and the last five classrooms were taught by teacher's having a bachelor's degree. The classroom means are reported in table 4. The

---

Insert table 4 here

---

average classroom achievement level of students taught by master's level teachers equalled 5.2 while the classroom average achievement taught by teachers having a bachelor's degree equalled 5.8. The computed F statistic equalled .686 and did not exceed the critical F statistic at the .05 level of significance ($F_{1,8,.05}=5.32$). The variability of the mean achievement scores can be tested by first transforming the classroom means using equation 1 and then calculating the ANOVA F-ratio using the transformed scores as the dependent measure. Table 4 presents the transformed scores for the classroom means. The analysis of variance F-ratio using these data equalled 5.803 which exceeded the critical test statistic at the .05 level of significance ($F_{1,8,.05}=5.32$). Thus the results here indicate that

teacher's with a bachelor's degree were less consistent than teacher's with a master's degree in terms of average classroom achievement.

Controlling classroom variability

When experimental units are not randomly assigned to treatment conditions, researchers often find an initial difference between comparison groups. As pointed out earlier researchers of teacher training effects have generally attempted to provide some adjustment for initial differences between comparison groups. Analysis of covariance provides one such adjustment procedure although the adjustment is often incomplete (Campbell and Erlebacher, 1970). While some adjustment for initial differences may be provided using ANCOVA, the groups may differ on other variables that have not been controlled and as a result strong causual relationships cannot be inferred even after controlling for initial differences in ability or previous achievement. Considering variability in this context, the initial heterogeneity of students within classrooms may be one alternative explanation for differences between classrooms taught by teachers having different training backgrounds. Suppose the data presented in Table 1 are a measure of initial achievement or ability for the 10 classrooms and the data reported in Table 5 are the results of a posttest administered at some later point in the school year. Again assuming the first five classrooms represent

Insert table 5 here

teachers having a master's degree and the last five classrooms being taught by teachers having a bachelor's degree, do the data support a difference in achievement between degree levels? Ignoring the initial scores the mean posttest score for the master's group equalled 8.4 and the bachelor's group mean equalled 11.0. The computed F ratio for the posttest data equalled 4.97 which is less than the critical F statistic at the .05 level of significance ($F_{1,8,.05}=5.32$). If the inital spelling scores are considered as a covariate in a single factor analysis of covariance the adjusted means equal 8.426 and 10.974 for the master's and bachelor's degree classrooms respectively. The computed F-ratio equals 4.09 which is also less than the critical F ratio at the .05 level of significance ($F_{1,7,.05}=5.59$). However if the researcher considers the initial within classroom variability as a covariate the adjusted means on the posttest equal 6.968 and 12.432 for master's and bachelor's degree classrooms respectively. The computed F-ratio equals 7.85 which is greater than the critical F statistic at the .05 level of significance ($F_{1,7,.05}=5.59$). Thus the difference is statistically significant and it may be concluded that classrooms taught by teachers having bachelor's degree achieved higher than

the classrooms taught by teachers having a master's degree. It would not be concluded however that the difference was caused by the degree level of the teacher. Since there were initial differences in classroom variability, there could be many other differences between the classrooms which could explain the difference in posttest scores besides the teachers' degree level that have not been controlled.

It is of course possible and possibly desirable to consider both initial achievement levels and initial classroom variability as covariates when comparing the posttest means. With the two covariates the adjusted means equalled 5.026 and 14.374 and the computed F-ratio equalled 72.53 which is much larger than the critical F statistic at the .05 level of significance ($F_{1,6,.05} = 5.99$). The present data set is unusual in that the initial variability and the initial mean achievement were not related yet both were related to the posttest scores. Furthermore the data set was created so that the lower the initial student variability the greater the gain in average achievement. Although the data are artificial they do provide a reasonable scenario and demonstrate how student variability might be used as a control variable.

Finally, if the researcher had considered consistency of classroom performance as an outcome of interest the researcher might consider comparing the variability of the posttest means between the two degree level teachers. Transforming the scores in Table 5 using equation 1 and

computing the analyis of variance comparing the mean transformed scores, the computed F ratio equalled .69 which is less th:n the critical F statistic at the .05 level of significance ($F_{1,8,.05}=5.32$). If the researcher had considered the initial classroom variability as a covariate the adjusted mean variances equalled 3.469 and 3.331 for the classrooms taught by master's and bachelor's degee level teachers respectively. The computed F-ratio equalled .00 which is again less than the critical F statistic ($F_{1,7,.05}=5.59$) and it would be concluded that there was insufficient evidence to indicate that the consistency of of posttest achievement differed betweeen the degree levels.

Discussion

The purpose of the present paper was to consider student variability within classrooms and the variability of average achievement across classrooms as outcomes of interest or possibly as a nuisance variable for researchers studying the effects of teacher training programs. Examining dispersion provides answers to interesting research questions that has not been examined in previous investigations. Since the development of an appropriate statistical method to answer these research questions is not well known, a major objective of this paper was to demonstrate the application of the technique developed by O'Brien (1978) which does provide a valid test for comparing variances regardless of the distribution form. The data set developed for this paper was limited and artificial but it

was sufficient for its purpose. What is needed now is actual data from classrooms and schools to investigate classroom variability. Do classrooms actually differ in student heterogeneity? What factors explain this variability? Is it the instructional philosophy of grouping students? Is it related to the instructional stategy adopted by the teacher and/or school? If classrooms do differ in their within group variability how does that affect teacher behaviors and the effectiveness of teacher behavior? These and many other questions can and should be answered by examining classroom variabilty. Examining classroom means can answer questions of interest to researchers but looking beyond the group means may also provide important information and increase our understanding of teacher training effects.

References

Ashton, P., Crocker, L. & Olejnik, S. (1986). Does teacher education make a difference? Final report submitted to the Institute on Student Assessment and Evaluation Grant No. 050-94640-850000, Florida Department of Education.

Bartlett, M.S. (1937). Proberties of sufficiency and statistical tests. Proceedings of the Royal Society, A901,160, 268-282.

Brown, B.W. & Saks, D.H. (1975). The production and distriution of cognitive skills within schools. Journal of Political Economy, 83, 571-593.

Campbell, D.T. & Erlebacker, A. (1970). How regresion artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.) Compensatory education :A national debate (Vol. 3) The disadvantage child. New York: Brunner/Mazel.

Cochran, W.G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. Annals of Eugenics, 11, 47-52.

Conover, W.J., Johnson, M.E. & Johnson, M.M. (1981). A comparative study of tests for homogeneity of variances with applications to the outer continental self bidding data. Technometrics, 23, 351-361.

Davis, C.R. (1964). Selected teaching-learning factors contributing to achievement in chemistyr and physics. Unpublished doctoral dissertation, University of North Carolina at Chapel Hill.

Hartley, H.O. (1950). The maximum F-ratio as a short-cut test for heterogeneity of variance. Biometrika, 37, 308-312.

Katzman, M.T. (1971). The political economy of urban schools. Cambridge, MA: Harvard University Press.

Murnane, R.J. (1975). The impact of scool resourdews on the learning of inner city children. Campridge, MA: Ballinger.

O'Brien, R.G. (1978). Robust techniques for testing heterogeneity of variance effects in factorial designs. Psychometrika, 43, 327-342,

O'Brien, R.G. (1981). A simple test for variance effects in experimental designs. Psychological Bulletin, 89, 570-574.

Olejnik, S. & Algina, J. (1987). Type I error rates and power estimates of selected parametric and nonparametric tests of scale. Journal of Educational Statistics, (in press)

Summers, A.A. & Wolfe, B.L. (1977). Do schools make a difference? American Economic Review, 67, 639-652.

Table 1

Observed scores, class means and variances for a hypothetical study

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10
	7	4	4	7	8	10	14	1	12	3
	4	4	6	7	4	4	12	1	10	6
	4	8	3	8	6	6	9	7	9	7
	5	7	2	6	6	2	7	1	6	9
	4	4	4	4	3	2	6	1	7	3
	6	8	3	4	3	1	6	1	4	2
	6	7	6	9	8	2	15	2	11	4
	4	6	7	3	3	3	7	8	3	8
	3	7	3	5	2	6	6	2	5	2
	7	5	2	7	7	4	8	6	13	6
mean	5	6	4	6	5	4	9	3	8	5
variance	2.00	2.67	3.11	3.78	5.11	7.33	11.78	8.00	12.22	6.44

Table 2

Transformed scores to test for variance equality

	Class 1	Class 2	Class 3	Class 4	Class 5
	4.599	4.557	-.194	.945	10.310
	1.056	4.557	4.530	.945	.862
	1.056	4.557	.987	4.488	.862
	-.125	1.014	4.530	-.236	.862
	1.056	4.557	-.194	4.488	4.405
	1.056	4.557	.987	4.488	4.405
	1.056	1.014	4.530	10.393	10.310
	1.056	-.167	10.435	10.393	4.405
	4.599	1.014	.987	.945	10.310
	4.599	1.014	4.530	.945	4.405
mean	2.00	2.67	3.11	3.78	5.11
variance	3.347	4.091	10.786	15.249	15.249

	Class 6	Class 7	Class 8	Class 9	Class 10
	42.058	28.789	4.224	18.132	4.322
	-.458	9.893	4.224	3.960	.779
	4.266	-.736	18.396	.417	4.322
	4.266	3.988	4.224	3.960	18.494
	4.266	9.893	4.224	.417	4.322
	10.171	9.893	4.224	18.132	10.227
	4.266	41.780	.681	9.865	.779
	.722	3.988	29.025	28.761	10.227
	4.266	9.893	.681	9.865	10.227
	-.458	.445	10.129	28.761	.779
mean	7.34	11.78	8.00	12.23	6.45
variance	158.446	178.901	81.454	115.920	33.102

Table 3

Analysis of variance summary table for effect of degree level on classroom variability.

Source	d.f	MS	F
Degree	1	848.18	19.99
Classroom: Degree	8	42.43	.688
Students: Classroom and Degree	90	61.69	

Table 4

Observed classroom means and means transformed for master and bachelor degree level teachers.

	Observed Classroom means		Means Transformed	
	Master	Bachelor	Master	Bachelor
	5	4	-.0583	3.6082
	6	9	.8166	13.8163
	4	3	1.9833	10.3164
	6	8	.8166	5.9415
	5	5	-.0583	-.1834
mean	5.2	5.8	.7	6.7
variance	.7	6.7	.706	30.309

Table 5

Observed posttest means and the means transformed for master and bachelor level teachers.

	Observed Posttest Means		Transformed Posttest Means	
	Master	Bachelor	Master	Bachelor
	11	12	9.411	1.0416
	10	11	3.017	-.4167
	7	9	2.142	5.4165
	8	10	-.483	1.0416
	6	13	7.683	5.4165
mean	8.4	11.0	4.3	2.5
variance	4.3	2.5	16.017	7.443